amcs

THE ROLE OF PARAMETER CONSTRAINTS IN EE AND OE METHODS FOR OPTIMAL IDENTIFICATION OF CONTINUOUS LTI MODELS

WITOLD BYRSKI, JĘDRZEJ BYRSKI

Department of Automatic Control
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Cracow, Poland
e-mail: wby@ia.agh.edu.pl, jbyrski@agh.edu.pl

The paper presents two methods used for the identification of Continuous-time Linear Time Invariant (CLTI) systems. In both methods the idea of using modulating functions and a convolution filter is exploited. It enables the proper transformation of a differential equation to an algebraic equation with the same parameters. Possible different normalizations of the model are strictly connected with different parameter constraints which have to be assumed for the nontrivial solution of the optimal identification problem. Different parameter constraints result in different quality of identification. A thorough discussion on the role of parameter constraints in the optimality of system identification is included. For time continuous systems, the Equation Error Method (EEM) is compared with the continuous version of the Output Error Method (OEM), which appears as a special sub-case of the EEM.

Keywords: continuous systems, parameter constraints in identification, modulating functions, transfer function normalization, least squares method.

1. Introduction

Continuous-time differential equations are the main mathematical tool for the description of many models of process physics. The knowledge of the parameters of these equations is necessary for simulation and control. One approach to the identification of a time continuous system given by the output differential equation

$$\sum_{i=0}^{n} a_i y^{(i)}(t) = \sum_{j=0}^{m} b_j u^{(j)}(t) \tag{1}$$

is its discretization, and identification of the discrete model. To have a continuous model, one has to convert the obtained parameters of the discrete model to the continuous version (1). Such indirect identification does not always guarantee proper results, especially when measurement disturbances occur. The discretization of the model (1) is also applied because identification methods for continuous time need information about derivative terms which are not available for measurement in most practical situations. Other approaches to identification of the model (1) when input or output data are given as non-uniformly sampled data can be found in works of Garnier and Wang (2008), Gillberg and Ljung (2009), Ljung and Wills (2010) as well as Johansson (2010).

The aim of this paper is to recall the background of a direct method for identification of continuous systems by using a modulating function with compact support and a convolution filter. Based on this methodology, a general and optimal identification method is presented as the Continuous Equation Error Method (CEEM). An important discussion about the role of the parameter constraints in the optimality of identification of time continuous systems is carried out. It will be shown that the assumption of different parameter constraints is equivalent to different normalizations of the model (1) and results in different quality of identification.

A suitable continuous version of the Output Error Method (COEM) is also derived. It is proved that the COEM can be treated as a special case of the CEEM. For these methods identification results are obtained and compared.

It is worth noting that the parameters a_n and b_m appearing in the model (1) are not normalized to one.

2. Method of modulating functions

In many cases of the modeling of Single Input-Single Output Continuous Linear Time Invariant (SISO CLTI) systems given by (1), only inputs $u = u^{(0)}$ and outputs

$y = y^{(0)}$ are available for direct measurements. The main problem in the identification of parameters a_i and b_j is caused by the unknown values of the derivatives $u^{(i)}$ and $y^{(i)}$ of input/output signals. Various filtration methods are to for obtain these derivatives, but all of them give only their approximations. Substituting these derivative approximations to the model (1) will cause the inequality of the left and right hand sides of the equation even under the assumption that all parameters are known while input u and output y measurements are perfect.

To guarantee the equality of both the sides of Eqn. (1) and to omit the problem of unknown derivatives, one may use special mathematical transformations which are based on integral convolution of both the sides of Eqn. (1) and a specially chosen function φ defined on a compact support (a finite window). This method was first used by Shinbrot (1957) and the functions φ were called modulating functions. In the works of Byrski and Fuksa (1995) as well as Byrski *et al.* (1999) one can find generalizations of this idea, which represents a continuous version of the equation error method. Below we briefly present this method.

Assume that the CLTI model of the continuous SISO system to be identified is of the form (1), where $m \leq n$, $y^{(i)}(t)$ and $u^{(i)}(t)$ are the i-th derivatives of the output and input, respectively, and the $n + m + 2$ unknown parameters a_i, b_j are constant. It can be assumed that perfect measurements of y and u on the interval $[t_0, T]$ are given. In order to avoid difficulties caused by the presence of derivatives in (1), this model can be transformed into a more convenient form by means of convolution. Choosing some special filtering modulating function φ with known derivatives $\varphi^{(i)}$, one can calculate the convolution of both the sides of the model (1) and the function φ. This function φ is supposed to be nonzero in the interval $[0, h]$, $m \ll n$ and zero outside this interval (φ is a function with compact support).

The convolution of two functions, e.g., $y^{(i)}(t)$ and $\varphi(t)$ (one of which is the i-th derivative), gives a new function $y_i(t)$, for $t \in [t_0 + h, T]$. From the well-known properties of integration by parts it also follows that the convolution of unknown derivatives of the output $y^{(i)}(t)$ or input $u^{(i)}(t)$ and the known function φ is equal to the convolution of a known output (input) signal $y(t)$, $(u(t))$ and a known derivative $\varphi^{(i)}$. Because of special properties of the assumed modulating function φ, the main formulas have the following form:

$$y_i(t)$$
$$\stackrel{def}{=} \left[y^{(i)} * \varphi \right](t) = \left[y * \varphi^{(i)} \right](t)$$
$$= \int_{-\infty}^{+\infty} y(\tau)\varphi^{(i)}(t - \tau) \, d\tau$$

$$= \int_{t-h}^{t} y(\tau)\varphi^{(i)}(t - \tau) \, d\tau = \int_{0}^{h} y(t - \tau)\varphi^{(i)}(\tau) \, d\tau.$$

$$u_j(t)$$
$$\stackrel{def}{=} \left[u^{(j)} * \varphi \right](t) = \left[u * \varphi^{(j)} \right](t)$$
$$= \int_{-\infty}^{+\infty} u(\tau)\varphi^{(j)}(t - \tau) \, d\tau$$
$$= \int_{t-h}^{t} u(\tau)\varphi^{(j)}(t - \tau) \, d\tau = \int_{0}^{h} u(t - \tau)\varphi^{(j)}(\tau) \, d\tau.$$

After integral transformation, the function $y^{(0)}(t) = y(t)$ is represented by $y_0(t)$ and, e.g., the first derivative $y^{(1)}(t)$ by $y_1(t)$. After convolution, the differential model (1) becomes an algebraic one with the sum of known functions y_i and u_i on both the sides of the equation,

$$\sum_{i=0}^{n} a_i y_i(t) = \sum_{j=0}^{m} b_j u_j(t). \tag{2}$$

These known functions are weighted by unknown parameters. The identification method should find the best parameters which give the least error between both the sides of (2). Equation (2) is well defined and valid for $t \in [t_0 + h, T]$ for any function φ of compact support (in this case $[0, h]$) and any distributions u, y from $\mathcal{D}'(\mathbb{R}^1)$ (Schwartz, 1966).

For the presented method, one should assume more about the function φ:

- The first assumption is connected with the support of the filtering function φ. It was assumed that the measurements of y and u are given on the finite time interval $[t_0, T]$, so the support of the function φ should also be finite and of width h less than $T - t_0$. Then the relation (2) will be fulfilled on $[t_0 + h, T]$. It is important that Eqn. (2) is independent of the initial conditions at the moment t_0 in (1). For different initial conditions there will be different $y^{(i)}(t)$ and $u^{(j)}(t)$, and hence functions $y_{i(t)}$ and $u_j(t)$, but they will all fulfil (2).

- If we assume that perfectly measured y and u are from space $L^2[t_0, T]$, then functions y_i, u_i obtained after the transformation should also be from the function space $L^2[t_0 + h, T]$, and all the left and right hand sides in relation (2) should be from the space $L^2[t_0 + h, T]$. This enables the use of the L^2 norm of the difference of both sides in (2) for identification purposes. Such a difference appears if the measurements of y and u are not perfect and are affected by noise. It will also happen if the order n of the model (1) is chosen incorrectly. Then the norm of

the equation error in the space $L^2[t_0 + h, T]$ will directly represent the quality of identification. Sufficient conditions assumed for function φ, namely, that its derivative $\varphi^{(n-1)}$ is continuous $\varphi \in C^{(n-1)}[0, h]$ and satisfies special boundary conditions,

$$\varphi^{(i)}(0) = \varphi^{(i)}(h) = 0, \quad i = 0, 1, 2, \ldots, n-1,$$

guarantee the equality $\left[y^{(i)} * \varphi\right](t) = \left[y * \varphi^{(i)}\right](t)$ in the procedure of integration by parts.

- The last assumption on the function φ is connected with the condition of uniqueness. This condition guarantees that from the equality $y * \varphi = 0$ it follows that $y = 0$ on the interval $[t_0 + h, T]$. This condition will eliminate, for example, any symmetric function φ: $\varphi(h/2 + t) = \varphi(h/2 - t)$, $t \in (0, h/2)$, which may give $y * \varphi = 0$ for nonzero y (e.g., $y = \sin(k2\pi t/h)$, and $k = 1, 2, 3, \ldots$). This problem was discussed in detail by Byrski and Fuksa (1995).

Various modulating functions were tested in the literature (Maletinsky, 1979; Co and Ydstie, 1990; Preisig and Rippin, 1993). We have found that the exponential modulating functions of the Loeb–Cahen type $\varphi(t) = t^M (h - t)^N$ are suitable for our purpose. Based on this modulating function method, a consistent and generalized theory, as well as various versions of optimal continuous identification algorithms, was presented by Byrski and Fuksa (1995; 1996; 1999; 2000). The main contribution of theze authors to this identification method is the assumption of a general type of constraints for parameters.

In this paper we will discuss the influence of the different constraints assumption (linear or quadratic normalizing conditions) on the identification quality. We also present a continuous version of the output error method.

3. Normalization of the transfer function

In fundamental monographs related to parameter identification, the parameters of the linear differential equation and those of the corresponding transfer function are always normalized.

The standard normalization of $G(s)$,

$$G(s) = \frac{b_m s^m + \cdots + b_1 s + b_0}{a_n s^n + \cdots + a_1 s + a_0}, \tag{3}$$

is concerned with the assumption $a_n = 1$ or $a_0 = 1$, (Young, 1981; Unbehauen and Rao, 1987). It is also applied to discrete systems (Eykhoff, 1974; Sinha and Kuszta, 1983; Soderstrom and Stoica, 1994).

Indeed, theoretically during input/output analysis for control purposes there is no difference between (3) and

two other transfer functions,

$$\begin{aligned}
G_1(s) &= \frac{(b_m/a_n)s^m + \cdots + (b_1/a_n)s + b_0/a_n}{s^n + \cdots + (a_1/a_n)s + a_0/a_n}, \\
G_2(s) &= \frac{(b_m/a_0)s^m + \cdots + (b_1/a_0)s + b_0/a_0}{(a_n/a_0)s^n + \cdots + (a_1/a_0)s + 1}.
\end{aligned} \tag{4}$$

However, in process identification, an arbitrary assumption $a_n = 1$ or $a_0 = 1$ leads to different results on identified parameter values and the value of the performance identification index. This problem for continuous systems was discussed by Byrski *et al.* (2003) and for the discrete systems by Yeredor (2006). In the following sections the complete results of investigations for continuous systems will be shown.

The basic form (3) is more useful for identification than the common form (4). For instance, in the case of an uncertain system structure and over-parametrization, there is a possibility to estimate the parameter a_n by zero as a result of the identification procedure.

We can formulate the following question: Is the identification of all parameters in (3) or (1) (without any constraints) possible so that the possible equation error will be minimized? The answer is positive but the best result will be trivial: all parameters equal zero. This is the reason why some extra constraint assumptions are needed, e.g., the parameter $a_n = 1$. This enables the formulation of the optimization task with constraints and obtaining a nontrivial solution for the rest of the parameters as well as a minimal value of the equation error. The assumption of parameter constraint $a_0 = 1$ will result in another optimal solution for the rest of the parameter values.

Hence one can state the problem of the general linear form for the parameter constraint,

$$\eta_1 a_0 + \cdots + \eta_{n+1} a_n + \eta_{n+2} b_0 + \cdots + \eta_{n+m+2} b_m = 1,$$

or in the vector form

$$[\eta_1, \eta_2, \ldots, \eta_{n+m+2}] \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \eta^T \Theta = 1,$$

where the vector of coefficients $\eta \in \mathbb{R}^{n+m+2}$ is arbitrarily chosen and $\eta \neq 0$. It represents the weights for different parameters a_i, b_j.

This condition eliminates the trivial solution $\Theta = 0$ and is a generalization of other linear constraints. For example, the vector $\eta = [0, 0, \ldots, 1, 0, \ldots, 0]^T$ gives the above mentioned common constraint for only one parameter $a_n = 1$, and the vector $\eta = [1, 0, \ldots, 0]^T$ will represent the constraint for only $a_0 = 1$. The assumption of the vector $\eta = [1, 1, \ldots, 1]^T$ means that the sum of all parameters a_i, b_j should be equal to 1. Then no parameter in (2) is preferred and the equation error is minimized by all the functions y_i and u_i in the best way.

Below we will derive a general formula for the identification of optimal parameters under linear constraints.

4. Equation error method for optimal identification of a continuous LTI system

Real continuous measurements of the input and output are given by $u, y \in L^2[t_0, T]$. For $y(t)$ and $u(t)$ the corresponding functions $y_i, u_j \in L^2[t_0 + h, T]$ are computed by the convolution procedure.

The term $\varepsilon \in L^2[t_0 + h, T]$ will denote the combined effects of unmeasurable noise or a general equation error,

$$\sum_{i=0}^{n} a_i y_i(t) = \sum_{j=0}^{m} b_j u_j(t) + \varepsilon(t). \qquad (5)$$

As the performance index for identification we can choose the norm of the difference of both the sides of the model (2),

$$J = \left\| \sum_{i=0}^{n} a_i y_i(t) - \sum_{j=0}^{m} b_j u_j(t) \right\|_{L^2[t_0+h,T]}. \qquad (6)$$

Denoting by $c(t) = [\widetilde{y}(t), \widetilde{u}(t)]$ the vectors of y_i and u_i after convolutions and by Θ the vectors of parameters $\Theta^T = [a, b] = [a_0, a_1, \ldots, a_n, b_0, \ldots, b_m]$, we obtain

$$\varepsilon(t) = c^T(t)\Theta = [\widetilde{y}(t), -\widetilde{u}(t)]\Theta$$
$$= [y_0(t), \ldots, y_n(t), -u_0(t), \ldots, -u_m(t)] \begin{bmatrix} a \\ b \end{bmatrix}. \qquad (7)$$

The statement of the minimization problem is

$$\min_{\Theta} J^2 = \min \|\varepsilon(t)\|_{L^2[t_0+h,T]}^2$$
$$= \min \left\| c^T(t)\Theta \right\|_{L^2[t_0+h,T]}^2,$$

and a general linear constraint (normalizing condition) is assumed as

$$\eta^T \begin{bmatrix} a \\ b \end{bmatrix} = \eta^T \Theta = 1, \qquad (8)$$

where the vector of weight coefficients $\eta \in \mathbb{R}^{n+m_2}$, $\eta \neq 0$ is chosen arbitrarily. The norm (6) has the form of an inner product in the space L^2:

$$J^2 = \left\langle c^T(t)\Theta, c^T(t)\Theta \right\rangle_{L^2[t_0+h,T]}$$
$$= \Theta^T \left\langle c(t), c^T(t) \right\rangle \Theta = \Theta^T G \Theta.$$

The real matrix G is a nonnegative definite, symmetric Gram $(n + m + 2) \times (n + m + 2)$ matrix of the inner products of functions which are elements of vector $c(t)$,

$$G = \begin{bmatrix}
\langle y_0, y_0 \rangle & \cdots & \langle y_0, y_n \rangle \\
\vdots & \ddots & \vdots \\
\langle y_n, y_0 \rangle & \cdots & \langle y_n, y_n \rangle \\
-\langle u_0, y_0 \rangle & \cdots & -\langle u_0, y_n \rangle \\
\vdots & \ddots & \vdots \\
-\langle u_m, y_0 \rangle & \cdots & -\langle u_m, y_n \rangle \\
& -\langle y_0, u_0 \rangle & \cdots & -\langle y_0, u_m \rangle \\
& \vdots & \ddots & \vdots \\
& -\langle y_n, u_0 \rangle & \cdots & -\langle y_n, u_m \rangle \\
& \langle u_0, u_0 \rangle & \cdots & \langle u_0, u_m \rangle \\
& \vdots & \ddots & \vdots \\
& \langle u_m, u_0 \rangle & \cdots & \langle u_m, u_m \rangle
\end{bmatrix}, \quad (9)$$

where

$$\langle y_i, u_j \rangle = \int_{t_0+h}^{T} y_i(\tau) u_j(\tau) \, d\tau$$
$$= \int_{t_0+h}^{T} \left[\int_{0}^{h} \left[\varphi^{(i)}(s) y(\tau - s) \right] ds \right.$$
$$\left. \cdot \int_{0}^{h} \left[\varphi^{(j)}(s) u(\tau - s) \right] ds \right] d\tau.$$

The Lagrange functional L for the above problem has the form

$$L = \Theta^T G \Theta + \lambda(\eta^T \Theta - 1). \qquad (10)$$

From the necessary condition for a minimum it follows that

$$\frac{\partial L}{\partial \Theta} = 2G\Theta + \lambda\eta = 0. \qquad (11)$$

The matrix G is nonsingular if there are uncorrelated disturbances in measurements or the rank of the model is less than that of the real system. From now on, if not otherwise stated, the matrix G will be assumed nonsingular. From (11) it follows that

$$\Theta = -0.5\lambda G^{-1}\eta, \qquad \eta^T \Theta = 1 = -0.5\eta^T \lambda G^{-1}\eta.$$

Hence, we get the final formula for the best identified parameters of the system (1)–(3),

$$\Theta^o = \frac{G^{-1}\eta}{\eta^T G^{-1}\eta}. \qquad (12)$$

Combining (12) and the formula for J we obtain the minimum of the performance index,

$$J^o = \sqrt{\Theta^{oT} G \Theta^o} = \frac{1}{\sqrt{\eta^T G^{-1}\eta}}. \qquad (13)$$

The introduction of the vector η also makes it possible to solve the problem in the more general case when the Gram matrix G is singular and its rank defect is equal to one or, generally, G is ill conditioned. Combining (8) and (11), we can use a special regularization technique,

$$[G^T, \eta] \cdot \begin{bmatrix} G \\ \eta^T \end{bmatrix} \Theta = [G^T, \eta] \cdot \begin{bmatrix} -0.5\lambda\eta \\ 1 \end{bmatrix}$$

and

$$[G^T G + \eta\eta^T] \Theta = [-0.5\lambda G^T\eta + \eta],$$
$$\Theta = [G^T G + \eta\eta^T]^{-1} [-0.5\lambda G^T\eta + \eta].$$

Hence

$$\eta^T\Theta = \eta^T [G^T G + \eta\eta^T]^{-1} [-0.5\lambda G^T\eta + \eta] = 1,$$
$$\lambda = 2\frac{\eta^T [G^T G + \eta\eta^T]^{-1}\eta - 1}{\eta^T [G^T G + \eta\eta^T]^{-1} G^T\eta}.$$

The matrix $[G^T G + \eta\eta^T]$ is nonsingular, and finally the following optimal parameter formula is obtained:

$$\Theta^o = [G^T G + \eta\eta^T]^{-1}$$
$$\cdot \left[\frac{1 - \eta^T [G^T G + \eta\eta^T]^{-1}\eta}{\eta^T [G^T G + \eta\eta^T]^{-1} G^T\eta} G^T\eta + \eta\right]. \quad (14)$$

If the matrix G is nonsingular, the above formula is reduced to (12). The vector η can be chosen almost arbitrarily (except that η is orthogonal to the kernel of G). In the next section it is shown that some η give better or worse results of identification if measurement noise occurs.

4.1. Optimal linear constraint in the EE method.
The optimal solution for the parameters (12) as well as the minimal value of the identification index (13) depend on the chosen vector η. Hence one can calculate a nontrivial vector η which additionally minimizes (13). To guarantee the existence of an optimal solution, one can formulate the minimization of the index (13) as a maximization of its denominator on a unit ball B. The vector $\eta \in B$ with unit norm $\eta^T\eta = 1$ will indicate the optimal direction in the space of parameters,

$$\min_B J^o \quad \Rightarrow \quad \max_B [\eta^T G^{-1}\eta] = \max_B J_1.$$

In this case the Lagrange function has the form

$$L = \eta^T G^{-1}\eta - \lambda(\eta^T\eta - 1),$$

and the optimality condition gives

$$\frac{\partial L}{\partial \eta} = 2G^{-1}\eta - 2\lambda\eta = 0 \quad \Rightarrow \quad G^{-1}\eta^o = \lambda\eta^o.$$

It can be seen that the *optimal vector η^o is an eigenvector* of the inverse Gram matrix G^{-1} and the Lagrange multiplier λ is its eigenvalue. Hence the value of the identification index is given by

$$J^o = \left(\sqrt{\eta^T G^{-1}\eta}\right)^{-1} = \left(\sqrt{\eta^T\lambda\eta}\right)^{-1} = \frac{1}{\sqrt{\lambda}}. \quad (15)$$

It is easy to see that the minimum of squared J is equal to the inverse of the maximal eigenvalue $\widehat{\lambda}_{max}$ of G^{-1} or to the minimal eigenvalue $\breve{\lambda}_{min}$ of G. It is clear that $\widehat{\lambda}_{max} = 1/\breve{\lambda}_{min}$,

$$\min_B J^o = \frac{1}{\sqrt{\widehat{\lambda}_{max}(G^{-1})}} = \sqrt{\breve{\lambda}_{min}(G)}.$$

The eigenvalues of the Gram matrix G are positive.

The optimal real vector η^o should thus be chosen as the eigenvector w_1 which corresponds to the minimal eigenvalue of G and the worst vector η is the eigenvector w_2 which corresponds to the maximal eigenvalue of G,

$$\eta^o = w_{min}(G) = w_1.$$

Then from (12) it follows that the normalized vector of optimal parameters Θ^o is also equal to the eigenvector $\Theta^o = w_1$,

$$\Theta^o = \frac{G^{-1}\eta}{\eta^T G^{-1}\eta} = \frac{\lambda w_1}{w_1^T \lambda w_1} = \frac{w_1}{w_1^T w_1} = w_1 = \eta^o. \quad (16)$$

The above result means that the norm of the parameter vector Θ is normalized to one, $\Theta^T\Theta = 1$.

This proves that the quality of the identification given by (13) depends on the chosen linear constraint.

In Fig. 1 one can see that in the two dimensional case the assumption of the only constraint $a_1 = 1$ results in a higher value of the quality index contour than the assumption of the quadratic constraint $a_0^2 + a_1^2 = 1$ (the unit circle).

Below, numerical tests are presented. Some of them were also presented by Byrski *et al.* (2003).

4.2. Numerical experiments.
Consider the transfer function $G(s)$ of a second order system,

$$G(s) = \frac{-0.8s + 2}{0.6s^2 + 2.3s + 1.5},$$

with the parameter vector

$$\Theta = [a_0, a_1, a_2, b_0, b_1]^T = [1.5, 2.3, 0.6, 2, -0.8].$$

Six cases of parameter normalization are tested.

1. Normalization of Θ to unit length yields $\Theta_1^T\Theta_1 = 1$,

$$G_1(s) = \frac{-0.226s + 0.565}{0.169s^2 + 0.649s + 0.423}.$$
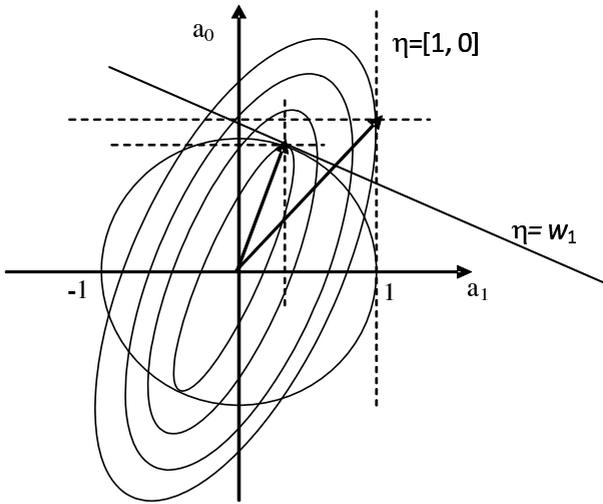
Fig. 1. Contours of the quality index and optimal solutions.

2. Normalization of Θ to $a_1 = 1$ gives Θ_2,

$$G_2(s) = \frac{-0.348s + 0.869}{0.261s^2 + s + 0.652}.$$

3. Normalization of Θ to $b_0 = 1$ gives Θ_3,

$$G_3(s) = \frac{-0.4s + 1}{0.3s^2 + 1.15s + 0.75}.$$

4. Normalization of Θ to $a_0 = 1$ gives Θ_4,

$$G_4(s) = \frac{-0.533s + 1.333}{0.4s^2 + 1.533s + 1}.$$

5. Normalization of Θ to $b_1 = 1$ gives Θ_5,

$$G_5(s) = \frac{s - 2.5}{-0.75s^2 - 2.875s - 1.875}.$$

6. Normalization of Θ to $a_2 = 1$ gives Θ_6,

$$G_6(s) = \frac{-1.333s + 3.333}{s^2 + 3.833s + 2.5}.$$

From experiments with the original transfer function $G(s)$, noisy input and output functions were measured and after convolution with the modulating function $\varphi(t) = t^5(h - t)^6$ (with support $h = 3$ and observation interval $T = 6$, noise variance $\delta = 0.01$) the algebraic model (2) was derived. Next, the Gram matrix (9) was calculated:

$$G = \begin{bmatrix} 6484.7 & 263.5 & -2287.4 & -3161.9 & 3195.9 \\ 263.5 & 2320.7 & -2.5 & -2860.5 & -259.6 \\ -2287.4 & -2.5 & 2323.5 & 361.7 & -1669.9 \\ -3161.9 & -2860.5 & 361.7 & 5263.6 & -352.7 \\ 3195.9 & -259.6 & -1669.9 & -352.7 & 3088.9 \end{bmatrix}.$$

The eigenvector w_1, which corresponds to the minimal eigenvalue of matrix G, is

$$\lambda_1 = 0.2695, \quad J = (\lambda_1)^{0.5} = 0.519,$$

$$w_1 = [0.429, 0.634, 0.163, 0.575, -0.236]^T.$$

The eigenvector w_2, which corresponds to the maximal eigenvalue of G, is

$$\lambda_2 = 11208, \quad J = (\lambda_2)^{0.5} = 105.86,$$

$$w_2 = [0.7187, 0.173, -0.272, -0.503, 0.355]^T.$$

The inverse corresponding to the above Gram matrix has the form

$$G^{-1}$$

$$= \begin{bmatrix} 0.6721 & 0.9930 & 0.2561 & 0.9010 & -0.3706 \\ 0.9930 & 1.4687 & 0.3785 & 1.3320 & -0.5473 \\ 0.2561 & 0.3785 & 0.0983 & 0.3434 & -0.1408 \\ 0.9010 & 1.3320 & 0.3434 & 1.2084 & -0.4966 \\ -0.3706 & -0.5473 & -0.1408 & -0.4966 & 0.2049 \end{bmatrix}.$$

Seven specially chosen vectors η will represent the standard transfer function normalizations. For these vectors the optimal parameters from (12) and the minimal value of the index J (13) will be derived. Note that all chosen vectors η will have unit length, and therefore, the comparison of different J is allowed.

Vectors η

$\eta_1 = w_1$ will yield the optimal and *the best* identification of parameters and results will be compared with G_1,

$\eta_2 = [0, 1, 0, 0, 0]^T$ will yield the optimal identification of parameters and results will be compared with G_2,

$\eta_3 = [0, 0, 0, 1, 0]^T$ will yield the optimal identification of parameters and results will be compared with G_3,

$\eta_4 = [1, 0, 0, 0, 0]^T$ will yield the optimal identification of parameters and results will be compared with G_4,

$\eta_5 = [0, 0, 0, 0, 1]^T$ will yield the optimal identification of parameters and results will be compared with G_5,

$\eta_6 = [0, 0, 1, 0, 0]^T$ will yield the optimal identification of parameters and results will be compared with G_6,

$\eta_7 = w_2$ will give optimal but *the worst* identification of parameters and results will be compared with G_1.

For the vectors $\eta_2, \ldots, \eta_6$ the calculation of the solution of optimal parameters $\Theta_2, \ldots, \Theta_6$ from (12) is very easy, as from the formula

$$\Theta^o = \frac{G^{-1}\eta}{\eta^T G^{-1}\eta}$$

it follows that, e.g., while assuming η as the vector $\eta_2 = [0, 1, 0, 0, 0]^T$ in the numerator of Θ there appears the second column of G^{-1} and in denominator we see the real number from the diagonal of G^{-1} (second column–second row). Adequately, assuming η as the vector $\eta_4 = [1, 0, 0, 0, 0]^T$, in the numerator the first column of G^{-1} appears and in denominator we have the real number from the diagonal of G^{-1} (first column-first row). Generally, for any standard normalization of the transfer function from G_2 to G_6, the solution of identification, is directly visible in the columns of G^{-1}. Hence so is the performance index J (13) which is equal to the inverse of the square root of a suitable diagonal number of G^{-1}. The final results of identification are presented in Table 1.

Table 1. Optimal parameters for different η_i.

η_i \ Θ	a_0	a_1	a_2	b_0	b_1	J_i
η_1	0.429	0.634	0.163	0.575	-0.236	0.519
η_2	0.676	1.000	0.257	0.906	-0.373	0.825
η_3	0.745	1.102	0.284	1.000	-0.411	0.909
η_4	1.000	1.477	0.381	1.340	-0.551	1.220
η_5	-1.809	-2.672	-0.687	-2.424	1.000	2.209
η_6	2.605	3.850	1.000	3.493	-1.432	3.189
w_2	0.719	0.173	-0.272	-0.503	0.355	105.8

For comparison of results and to see the differences, let us normalize the parameters in Case 1 to $a_2 = 1$. From

Table 2. Normalized optimal parameters.

η_i \ Θ	a_0	a_1	a_2	b_0	b_1
W_1	2.889	3.890	1.000	3.527	-1.447
η_6	2.605	3.850	1.000	3.493	-1.432

Table 1 we can see that for different vectors η different parameters Θ are identified. The last column presents the different values of the performance index (13), which represents the equation error for each chosen η,

$$\min_{\Theta} J = \min_{\Theta} \sqrt{\int_h^T \varepsilon^2(t)\, \mathrm{d}t} = \left[\sqrt{\eta^T G^{-1} \eta} \right]^{-1}.$$

The minimal value of J is given by the minimal eigenvector w_1 (first row of Table 1). The other normalization with unit parameter $a_i = 1$ gives worse results. The worst case is for the maximal eigenvector w_2 (last row of Table 1).

It is easy to compare the obtained results with perfect theoretical results visible in transfer functions G_1 to G_6.

Extra experiments. For each identified model one can calculate the value of another performance index based on the least squared differences between the identified parameters and real parameters of normalized models G_j (parameters error),

$$J_j = \sqrt{\sum_{i=1}^{5} \left(p_{ir}^j - p_{im}^j \right)^2}, \quad j = 1, 2, \dots, 6,$$

where p_{ir}^j is the i-th real parameter of the G_j transfer function and p_{im}^j is the i-th identified parameter of the j model from Table 1.

The results of the parameter error are given in Table 3. Table 3 also confirms the different quality of iden-

Table 3. Parameter error.

η_j	w_1	η_2	η_3	η_4	η_5	η_6	w_2
J_j	0.022	0.051	0.052	0.062	0.235	0.216	1.410

tification J_j for different η_i. From Table 1 (equation error) and 3 (parameter error) it follows that the very common assumption of constraint η_6 ($a_n = 1$) gives one of the worst results. Hence, the model form like $G_6(s)$ should not be used for identification purposes; however, this model is almost always used in the discrete type of identification (Gauss method of LSI).

For testing the quality of the proposed optimal identification solutions (with optimal constraint $\eta = w_1$), other six experiments were conducted with a different level of measurement noise variance δ^2. Six different Gram matrices G will give six different sets of optimal parameters. The results of experiments and the index J (13) are given in Table 4.

One can see that the same high level of error $J \cong 3$ occurs for the vector η_6 under low noise $\delta^2 = 0.01$ (used in the experiments in Table 1) and optimal vector η_1 under as many as ten times greater noise $\delta^2 = 0.1$ (Table 4).

Table 4. Optimal parameters $\Theta = w_1$ and noisy measurements.

δ^2	a_0	a_1	a_2	b_0	b_1	J
0	0.424	0.649	0.167	0.566	-0.225	0.027
0.01	0.429	0.634	0.163	0.575	-0.236	0.519
0.05	0.424	0.668	0.165	0.560	-0.180	2.780
0.1	0.410	0.649	0.205	0.576	-0.190	3.383
0.5	0.531	0.412	0.121	0.500	-0.533	6.218
1	0.472	0.514	0.131	0.633	-0.306	7.445

Conclusion. For the linear continuous system the use of the Equation Error Method (EEM) and the optimal choice of the linear constraint for the parameters guarantee much better results of parameter identification than other standard normalizations.

5. Output error method for identification of a continuous LTI system

In this section we will derive formulas which follow from the output error method, and we will prove that the same results can be obtained from the above discussed general EE method as its special case (η_4) for a special coefficient vector $\eta = [1, 0, \ldots, 0]^T$, which represents the commonly assumed constraint for only one parameter $a_0 = 1$.

Let us consider the LTI system with the parameter $a_0 = 1$,

$$G(s) = \frac{b_m s^m + \cdots + b_1 s + b_0}{a_n s^n + \cdots + a_1 s + 1}, \tag{17}$$

and the corresponding output model differential equation,

$$y(t) + \sum_{i=1}^{n} a_i y^{(i)}(t) = \sum_{j=0}^{m} b_j u^{(j)}(t).$$

Minimization of the error only between the real output measurements $y(t)$ and the rest of the model and not the whole sides of the equation is the substance of the output error method,

$$\varepsilon_r(t) = y(t) - \left[-\sum_{i=1}^{n} a_i y^{(i)}(t) + \sum_{j=0}^{m} b_j u^{(j)}(t) \right]. \tag{18}$$

Having used the convolution transformation with a proper function φ in (18), we obtain

$$\varepsilon(t) = y_0(t) - \left[-\sum_{i=1}^{n} a_i y_i(t) + \sum_{j=0}^{m} b_j u_j(t) \right].$$

We assume the performance index of identification as the norm of the output error function,

$$J = \|\varepsilon\|$$
$$= \left\| y_0(t) - \left(-\sum_{i=1}^{n} a_i y_i(t) + \sum_{j=0}^{m} b_j u_j(t) \right) \right\|_{L^2[t_o+h,T]}. \tag{19}$$

Denoting by $c(t)$ the shortened vector of convolutions $\tilde{y}(t)$ and $\tilde{u}(t)$ and by Θ the shortened vectors of parameters

$$\Theta^T = [a, b] = [a_1, \ldots, a_n, b_0, \ldots, b_m],$$

we get

$$\varepsilon(t) = y_0(t) - c^T(t)\Theta$$
$$= y_0(t) - [-\tilde{y}(t), \tilde{u}(t)]\Theta = y_0(t)$$
$$- [-y_1(t), \ldots, -y_n(t), u_0(t), \ldots, u_m(t)] \begin{bmatrix} a \\ b \end{bmatrix}.$$

The statement of the minimization problem takes the form ($t_0 = 0$):

$$\min_{\Theta} J^2 = \min_{\Theta} \|\varepsilon(t)\|_{L^2[h,T]}^2$$
$$= \min_{\Theta} \left\| y_0(t) - c^T(t)\Theta \right\|_{L^2[h,T]}^2,$$

where

$$J^2 = \langle y_0(t) - c^T(t)\Theta, y_0(t) - c^T(t)\Theta \rangle$$
$$= \langle y_0(t), y_0(t) \rangle - 2 \langle c^T(t)\Theta, y_0(t) \rangle + \Theta^T G\Theta.$$

There exists an optimal and nontrivial solution for Θ without any extra constraint for the parameters,

$$G = \begin{bmatrix} \langle y_1, y_1 \rangle & \cdots & \langle y_1, y_n \rangle \\ \vdots & \ddots & \vdots \\ \langle y_n, y_1 \rangle & \cdots & \langle y_n, y_n \rangle \\ -\langle u_0, y_1 \rangle & \cdots & -\langle u_0, y_n \rangle \\ \vdots & \ddots & \vdots \\ -\langle u_m, y_1 \rangle & \cdots & -\langle u_m, y_n \rangle \\ & -\langle y_1, u_0 \rangle & \cdots & -\langle y_1, u_m \rangle \\ & \vdots & \ddots & \vdots \\ & -\langle y_n, u_0 \rangle & \cdots & -\langle y_n, u_m \rangle \\ & \langle u_0, u_0 \rangle & \cdots & \langle u_0, u_m \rangle \\ & \vdots & \ddots & \vdots \\ & \langle u_m, u_0 \rangle & \cdots & \langle u_m, u_m \rangle \end{bmatrix}. \tag{20}$$

The real matrix G is a symmetric Gram $(n + m + 1) \times (n + m + 1)$ matrix of the inner products of functions which are elements of the vector $c(t)$. From the necessary condition for a minimum it follows that

$$\frac{\partial J^2}{\partial \Theta} = -2 \langle c(t), y_0(t) \rangle + 2G\Theta = 0,$$

where

$$\Theta^o = G^{-1} \langle c(t), y_0(t) \rangle = G^{-1} \begin{bmatrix} -\langle y_1(t), y_0(t) \rangle \\ \vdots \\ \langle u_m(t), y_0(t) \rangle \end{bmatrix}$$
$$= G^{-1} \begin{bmatrix} -\int_h^T [y_1(t)y_0(t)] \, dt \\ \vdots \\ \int_h^T [u_m(t)y_0(t)] \, dt \end{bmatrix}. \tag{21}$$

This is a continuous version of the Least Squares (LS) estimator for optimal parameter identification. A suitable version is well known for discrete systems,

$$\Theta^o = [C^T C]^{-1} C^T Y.$$

Here the vector Y of discrete measurements of $y(t)$ represents the continuously filtered output function $y_0(t)$ in (21) and the vector inner product $C^T Y$ is substituted by its continuous version. The regression matrix $C^T C$ is a discrete version of the Gram matrix G.

As mentioned above, the same result as in the formula (21) can be obtained directly from the general equation error method and the formula (12) after the assumption of a specific vector η, $\eta = [1, 0, 0, \ldots, 0]$. In such a sense the EE method with a general linear constraint and an arbitrarily chosen vector η is the most general method for identification which also includes different output methods. For instance, assuming $\eta = [0, 1, 0, \ldots, 0]$ ($a_1 = 1$) means that error minimization will pertain to the best fit of the first derivatives only, i.e., to the difference between the first derivative of output $y^{(1)}(t)$ taken from continuous measurements and the first derivative of output obtained from the model. Of course, if such a measurement of the first derivative is impossible, it will concern the function $y_1(t)$ obtained after the convolution transformation of $y(t)$.

6. Conclusions

In the paper the method of convolution transformation of the linear continuous dynamical model was presented. After the transformation an algebraic model is obtained which is suitable for identification. This model is valid for any initial conditions. However, for different initial conditions the different new functions $y_i(t)$ and $u_j(t)$ are generated and hence different Gram matrices are obtained, while the solution (12) will always give the optimal parameters to best satisfy the model (1).

The important role of the proper choice of parameter constraints in the optimal identification procedure was explained. It was shown that the common assumption of the parameter $a_n = 1$ or $a_0 = 1$ and the standard form of the transfer function, e.g., (4), is not always optimal for the identification procedure. Along with the handling of this problem, a discussion of connections between two different approaches to identification is performed in the paper. The first approach is based on the analysis of the equation error and the second on the analysis of the output error method. A proper choice of the normalization vector η enables the minimization of the error between any of the measurable state variables from the real system and from the model or any of their linear combinations. The new version of continuous LS method for output error with the use of modulating functions was presented.

In the publication by Byrski and Fuksa (1996), the EE type of filter was used in an on-line identification procedure which co-operates with an adaptive stabilizing LQ state controller. State stabilizing control is proportional to the transformed state $x(t) = f[y_i(t)]$ concurrently generated by the parameter identifier. Hence such a type of identifier can work as a special type of state observer. Numerical experiments showed good performance of this kind of linear adaptive controller. In the paper by Byrski and Fuksa (2001), the mathematical proof of the asymptotic stability of such simultaneous co-operation of the state and parameters estimators and feedback control was presented.

Acknowledgment

References

Byrski, W. and Fuksa, S. (1995). Optimal identification of continuous systems in ℓ_2 space by the use of compact support filter, *International Journal of Modelling & Simulation, IASTED* **15**(4): 125–131.

Byrski, W. and Fuksa, S. (1996). Linear adaptive controller for continuous system with convolution filter, *Proceedings of the 13th IFAC Triennial World Congress, San Francisco, CA, USA*, pp. 379–384.

Byrski, W. and Fuksa, S. (1999). Time variable gram matrix eigen-problem and its application to optimal identification of continuous systems, *Proceedings of the European Control Conference ECC99, Karlsruhe, Germany*, F0256.

Byrski, W. and Fuksa, S. (2000). Optimal identification of continuous systems and a new fast algorithm for on line mode, *Proceedings of the International Conference on System Identification, SYSID2000, Santa Barbara, CA, USA*, PM 2–5.

Byrski, W. and Fuksa, S. (2001). Stability analysis of CLTI state feedback system with simultaneous state and parameter identification, *Proceedings of the IASTED International Conference on Applied Simulation and Modelling, ASM01, 2001, Marbella, Spain*, pp. 7–12.

Byrski, W., Fuksa, S. and Byrski, J. (1999). A fast algorithm for the eigenproblem in on-line continuous model identification, *Proceedings of the 18 IASTED International Conference on Modelling, Identification and Control, MIC99, Innsbruck, Austria*, pp. 22–24.

Byrski, W., Fuksa, S. and Nowak, M. (2003). The quality of identification for different normalizations of continuous transfer functions, *Proceedings of the 22 IASTED International Conference on Modelling, Identification and Control, MIC03, 2003, Innsbruck, Austria*, pp. 96–101.

Co, T. and Ydstie, B. (1990). System identification using modulating functions and fast Fourier transforms, *Computers & Chemical Engineering* **14**(10): 1051–1066.

Eykhoff, P. (1974). *System Identification, Parameter and State Estimation*, J. Wiley, London.

Garnier, H. and Wang, L. (Eds.) (2008). *Identification of Continuous-time Models from Sampled Data*, Advances in Industrial Control, Springer-Verlag, London.

Gillberg, J. and Ljung, L. (2009). Frequency domain identification of continuous time ARMA models from sampled data, *Automatica* **4**(6): 1371–1378.

Johansson, R. (2010). Continuous-time model identification and state estimation using non-uniformly sampled data, *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems, MTNS 2010, Budapest, Hungary*, pp. 347–354.

Ljung, L. and Wills, A. (2010). Issues in sampling and estimating continuous-time models with stochastic disturbances, *Automatica* **46**(5): 925–931.

Maletinsky, V. (1979). Identification of continuous dynamical systems with spline-type modulating functions method, *Proceedings of the IFAC Symposium on Identification and SPE, Darmstadt, Germany*, Vol. 1, p. 275.

Preisig, H.A. and Rippin, D.W.T. (1993). Theory and application of the modulating function method, *Computers & Chemical Engineering* **17**(1): 1–16.

Schwartz, L. (1966). *Théorie des distributions*, Hermann, Paris.

Shinbrot, M. (1957). On the analysis of linear and nonlinear systems, *Transactions of the American Society of Mechanical Engineers, Journal of Basic Engineering* **79**: 547–552.

Sinha, N.K. and Kuszta, B. (1983). *Modeling and Identification of Dynamic Systems*, Van Nostrand RC, New York, NY.

Soderstrom, T. and Stoica, P. (1994). *System Identification*, Prentice Hall, London.

Unbehauen, H. and Rao, G. P. (1987). *Identification of Continuous Systems*, North-Holland, Amsterdam.

Yeredor, A. (2006). On the role of constraints in system identification, *4th International Workshop on Total Least Squares and Errors-in-variables Modelling, Leuven, Belgium*, (see also http://diag.mchtr.pw.edu.pl/damadics).

Young, P. (1981). Parameter estimation for continuous-time models—A survey, *Automatica* **17**(1): 23–39.

Witold Byrski obtained the Ph.D. degree in automatic control in 1976 from the AGH University of Science and Technology in Cracow, Poland. He received his *doctor habilitatus* (D.Sc.) degree in 1994 and the professorial title in 2009. Presently he is employed as a professor in the Process Group at the Department of Automatic Control, Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, AGH-UST in Cracow. Since 1998 he has been the deputy director of the Department of Automatic Control. His field of research includes new integral algorithms for exact state observation and convolution methods for identification of continuous dynamic models of high order. Results of his research have been published in a monograph, books, chapters papers, international conferences and congresses (over 90 reviewed publications and over 100 other research reports).

Jędrzej Byrski obtained the M.Sc. degree in computer science in 1997. Presently he is employed as an assistant in the Real-time Computer Systems Group at the Department of Automatic Control, AGH-UST in Cracow. His field of research covers safety problems in computer networks and firewall software, the CORBA system, real time computer systems in the QNX operating system, and mathematical algorithms for identification and fault detection. He has participated in many scientific research projects funded by the Polish State Committee for Scientific Research as well as European research projects (e.g., at the University of Birmingham, UK, designing the interface for CLP Prolog—a package for operational control of water systems), and has worked on the development of software for process decision optimization.